Symptom cluster patterns based on text clustering method of COVID-19 and its population characteristics in Sichuan province, China

Xiuwei Cheng
Center for Disease Control and Prevention of Sichuan

Hongli Wan
Sichuan University

Heng Yuan
Center for Disease Control and Prevention of Sichuan

Lijun Zhou
Center for Disease Control and Prevention of Sichuan

Chongkun Xiao
Center for Disease Control and Prevention of Sichuan

Suling Mao
Center for Disease Control and Prevention of Sichuan

Zhirui Li
Center for Disease Control and Prevention of Sichuan

Fengmiao Hu
Center for Disease Control and Prevention of Sichuan

Chuan Yang
Center for Disease Control and County of Anyue County

Wenhui Zhu
Sichuan University

Jiushun Zhou
Center for Disease Control and Prevention of Sichuan

Tao Zhang (✉ statzhangtao@scu.edu.cn)
Sichuan University

Abstract

Background: Evidence of the aggregation of COVID-19 symptoms was still limited. Exploration of likely cluster patterns of symptoms may be helpful.

Methods: This study enrolled a total of 1067 COVID-19 cases. Symptom cluster patterns were explored by text clustering method. A multinomial logistic regression was applied to revealed its population characteristics. And time intervals between symptoms onset and the first visit was analyzed to take into account the symptoms progression over time.

Results: Based on text clustering, the symptoms were summarized into 4 groups. Except for the group without obvious symptoms, the dominate individual symptoms under the other three groups were fever (68.7%), expectoration (59.4%) and fatigue (42.7%), respectively. And the most frequent symptom combinations were fever only (47.8%), expectoration only(19.8%) and fatigue accompanied with fever (4.2%), respectively. People aged 45-64 years were more likely to have symptom group 4 than those aged 65 years or older(OR = 2.66, P = 0.015) and had longer time intervals.

Conclusions: The symptoms cluster patterns and common symptom combinations under each pattern may provide information for identification of the infected. Middle-aged population was a group expecting more attention, from the perspective of medical delays.

Background

The coronavirus Disease-19 (COVID-19) pneumonia has caused significant morbidity and mortality worldwide[1]. At this time, while the epidemic has been under well control to a certain degree, there are possibilities of re-outbreak caused by sporadic cases[2]. Warning signs from early symptoms may be helpful.

Clinical symptoms, as an indicator for identification and diagnosis, plays a vital role in the early detection and treatment. COVID-19 has a wide range of clinical manifestations, ranging from asymptomatic to severe viral pneumonia [3, 4]. It has been widely confirmed that fever, dry cough, expectoration, and fatigue were the most common symptoms of COVID-19 patients[3, 5–7]. In addition, symptoms of cardiovascular system[8], digestive system[9], skin[10], and loss of taste and smell[11] were also reported. A number of studies have also conducted dynamic studies of the clinical course of symptoms[5, 12, 13]. There are also studies having revealed similarities and differences in symptoms between COVID-19 and other respiratory infections like influenza.

Nevertheless, most of published studies have primarily focused on descriptions of individual symptoms. Noting that two or more symptoms normally coexisted in one infected person and there are usually similarities of symptom combinations between individuals, the purpose of this study was to explore whether there are likely cluster patterns of symptoms in COVID-19 patients basing on the aggregation of symptoms with text clustering method. Giving that there are both overlap and variances in symptoms of

COVID-19 and other infectious diseases like flu[14], the likely cluster patterns and common symptom combinations under each pattern may provide some enlightenment to improve early identification of COVID-19. Based on the exploration of symptom cluster patterns, this study also analyzed the population characteristics of different symptoms groups. Furthermore, to take into account the progression of symptoms over time, time intervals between symptoms onset and the first visit were analyzed.

Methods

Study design and data source

In this cohort study, a total of 1067 laboratory confirmed cases of COVID-19 from January 21, 2020 to November 20, 2020 in Sichuan Province were included. All data were sourced via the Epidemic Registration System of the Sichuan Center for Disease Control and Prevention (CDC). This study was approved by Ethics Committee of Sichuan Provincial Center for Disease Control and Prevention (NO. SCCDCIRB 2020-007). Written informed consent was obtained from each of the subjects.

Statistical analysis

Symptoms were firstly explored by text clustering method on the basis of Euclidean distance. Then combined with pathophysiology[15, 16] and consultation with clinical experts from the Sichuan Provincial Center for Disease Control and prevention, the symptoms were summarized into different groups. Bar charts were used to give a visual representation of the symptom combinations under each group. In addition, a word cloud map was used to display the dominating symptoms of different groups according to their frequencies. Categorical variables were described by frequency and percentage, continuous variables were described by median and interquartile ranges (IQR).

Based on the classification, a multinomial logistic regression was applied to identify potential factors associated with the symptom groups. Symptoms of group 1 was the reference category in the multinomial regression model. Additionally, the interval between symptoms onset and the first visit in each group was represented by the change of proportions of each group over time.

Figure 1 is a diagram showing the procedure of our analysis. In this article, the text clustering was conducted with Python version 3.7.6 and the rest statistical analyses were conducted with R version 4.0.3. P value less than 0.05 was considered statistically significant.

Results

Symptom cluster patterns

As of Nov. 20, 2020, information of 1067 cases were collected. Based on the results of text clustering, symptoms were summarized into the following 4 groups: *Group 1: No-obvious symptoms,* refers to those with no obvious symptoms but positive nucleic acid test; *Group 2: Mainly fever and/or dry cough*, refers to those with fever as the main symptoms, or accompanied by dry cough; *Group 3: Mainly upper respiratory*

tract infection symptoms, refers to those mainly with expectoration and upper respiratory tract infection symptoms, such as, pharyngodynia, stuffy nose and runny nose, or accompanied by fever; *Group 4: Mainly cardiopulmonary, systemic and/or gastrointestinal symptoms*, refers to those whose main symptoms were cardiopulmonary symptoms such as shortness of breath, dyspnea, chest tightness, chest pain, and/or systemic symptoms such as fatigue, chills and muscle aches, and/or symptoms of the gastrointestinal system such as nausea, vomiting and diarrhea, sometimes accompanied by fever and upper respiratory tract symptoms.

The results showed that more than half (50.7%) of the infected persons did not show obvious symptoms at the first visit, i.e. in the group 1. For the three groups with obvious symptoms, the proportions were 12.6%, 10.0% and 26.8%, respectively. Group 4, that were cardiopulmonary, systemic and/or gastrointestinal symptoms had higher proportion.

In order to profile the symptoms composition under each group, bar charts were applied to visualize the particular symptoms under each group (Fig. 2). It could be seen that there were overlaps and interactions of symptoms under a same group. In symptom group 1, all with no-obvious symptoms (541 cases,100%); In symptom group 2, the most frequent symptom combinations were fever only (64 cases, 47.8%) , followed by dry cough only(42 cases, 31.3%); In symptom group 3, the most frequent symptom combinations was expectoration only(21 cases, 19.8%), followed by fever complicated with expectoration (10 cases, 9.4%); In symptom group 4, the most frequent symptom combinations was fatigue complicated with fever (12 cases, 4.2%), the incidence of headache complicated with fever was also high (11 cases, 3.8%).

Fig. 3 shows a word cloud based on the frequency of individual symptoms under each group. The larger the font, the higher the frequency. It could be seen that fever and dry cough were the two most frequent symptoms in general, with frequencies of 64.4% and 38.8%, respectively, followed by expectoration (12.0%) and fatigue (11.4%). Fever (68.7%) and dry cough(52.24%) were the dominant symptoms in group 2; Expectoration(59.4%) was the dominant symptom in group 3; And fatigue(42.7%) and headache(26.2%) were the dominant symptoms in symptom group 4. Under each symptom group, symptoms showed some clustering around the dominant symptoms.

Population distribution in different symptom groups

Of the 1067 cases, 699(65.51%) were males and 368(34.49%) were females. The median age was 35 years. Population characteristics of the above symptom groups were summarized in Table 1. According to Sijia Tian et al.[17], we divide the infected into four groups aged 0-12,13-44,45-64, ≥65 years. The majority of infected persons were in 13-44(613 cases, 57.45%) and 45-64(344 cases, 32.23%) years age groups. For comorbidities, the prevalence of hypertension was 6.84%, while were 2.44%, 1.88% and 2.06% of diabetes, lung disease and cardiovascular disease, respectively. In addition, 41.24% of the infected patients were imported cases and 26.43% were infected with family cluster.

Factors associated with different symptom groups

The results of multinomial logistic regression (Table 2) showed that age, comorbidities and epidemiological characteristics were all independent influencing factors of the presence of symptom group 4, namely symptoms such as cardiopulmonary, systemic and/or gastrointestinal symptoms. Compared with the 0-12 years age group, the odds of symptoms of group 4 increased in both the 13-44 years and 45-64 years age groups (OR = 4.08, P = 0.032; OR = 5.91, P = 0.007). In addition, in order to further analyze whether there is any difference in the odds of the three obvious symptom groups in age groups compared with ≥65 years group, an analysis was carried out with the group of ≥65 years as the reference. It is noteworthy that people aged 45-64 years were more likely to develop symptoms of group 4 (OR = 2.66, P=0.015) when compared with the ≥65 years group.

No significant differences in the odds of the above three obvious symptom groups were detected between the sexes. For the comorbidities, the odds of showing symptom group 2 was no significant differences between patients with and without diabetes (P = 0.111), but in those with diabetes, the odds of group 3 and group 4 had significantly escalated (OR = 29.43, P= 0.004; OR = 41.72, P = 0.001), indicating diabetes a strong risk factor for upper respiratory tract symptoms, cardiopulmonary, systemic and/or gastrointestinal symptoms. In addition, the results showed that there was no significant difference in the odds of all the three obvious symptom groups between patients with or without hypertension, lung disease or cardiovascular disease.

Besides, the results showed that, the incidences of all the 3 obvious symptom groups were lower in the imported cases and the patients infected with cluster than in the indigenous cases and non-clustered cases, respectively (OR< 1, P< 0.05).

Time intervals between symptoms onset and the first visit

In all the symptomatic cases, the median time interval between symptoms onset and the first visit was 1 day, and the interquartile range was (0,3) days. 47.5% of symptomatic patients visited a medical institution on the day of symptoms onset, 15.4% one day after onset and 11.4% two days after onset, and 25.7% sought medical treatment three days or more after onset. Fig. 4 is a graph displaying the proportions of cases with the three groups of obvious symptoms as the time intervals lengthened. It could be seen that the proportion of symptom group 2 was decreasing as the time interval lengthened, while in symptom group 4, it was increasing over longer time intervals, and in symptom group 3, its proportion peak was in the middle.

Noting that the results showed people aged 45-64 years were more likely to show more severe symptoms than people aged 65 or older, in order to explore whether this was affected by the progression of symptoms, we analyzed the time intervals between symptom onset and visit in different age groups. The results showed that the median time intervals of 0-12, 13-44, 45-64 years groups were all 1 day, while it was 0 day of ≥65 years group (Fig. 5). And the ranges were longer in 13-44 years age group and 45-64 years age group, with ranges of (0,14) days and (0,15) days respectively, while which in 0-12 years age group and ≥65 years age group were (0,7) days and (0,8) days respectively. Patients aged 12-64 years had longer time intervals.

Discussion

This study focused on the aggregation of different symptoms, have explored the symptoms cluster patterns. Like many prior studies, we found fever, dry cough were the most common symptoms, followed by expectoration and fatigue[18, 19].There existed probable different patterns of symptoms, which could be summarized into four groups. And we illustrated the specific symptom combinations under each group. The most frequent individual symptom were fever, expectoration and fatigue in the three groups with obvious symptoms, respectively. And the most frequent symptom combinations were fever only, expectoration only and fatigue accompanied with fever, respectively. It had been confirmed that both COVID-19 and influenza have fever, cough, expectoration and fatigue with their main symptoms[20–22]. On the other hand, some symptoms such as vomiting, nasal congestion, runny nose and ocular symptoms are more common in influenza than COVID-19[21–23], while in COVID-19, symptoms such as fatigue, neurological symptoms(like headache), gastrointestinal symptoms are more common[22, 24, 25].Giving there are both overlaps and variation between the two, judgment relying on a single symptom is likely to bring misunderstanding, which is of little value for early identification. Therefore, awareness of the combination of symptoms of COVID-19 and commonly accompanying symptoms may provide some information for distinguishing it from other respiratory infections like influenza through symptoms.

Besides, the results of multinomial logistic regression showed that compared with younger aged (0–12 years) people, those aged 13–44,45–64 and ≥ 65 years had increased odds of developing symptom group 4. This has been confirmed in previous studies that immunosenescence and inflamm-aging may be the origin[26, 27]. For the comorbidities, patients with chronic diseases such as diabetes were more likely to show symptoms of group 4, which was also been confirmed[28]. In addition, the results showed that for the imported cases and the clustered cases, the odds of symptom group 2, group3 and group 4 were all lower than indigenous cases and non-cluster cases, respectively. For imported cases, the entry quarantine of the imported may provide an explanation. And for the results that cases with non-cluster had more severe symptoms, this may be reasonable that infecting occurred within a same family, work unit, nursery or school means an infected person was more likely to be found as a close contact of whom with which he/she was clustered, thus was more likely to be found at the early stage [26].

For the result that the prevalence of symptom group 4 (26.8%) was higher than that of group 2(12.6%) and group 3 (10.0%), we took consideration of the progression of symptoms over time. From the results of the time intervals analysis, the proportion of symptom group 2 decreased with the extension of the time interval, on the other hand, the proportion of group 4 increased. This suggested a chronological order of occurrence of different symptom patterns, which had also been confirmed in some previous studies examining the dynamic changes of symptoms. According to Joseph R. et al.[29], whose analysis of the symptoms in 55,924 confirmed cases based on a Markov-process showed that there was a possible order in the development of COVID-19 symptoms. The symptoms may progress initially with fever or cough followed by upper respiratory symptoms such as sore throat, after fatigue and other systemic symptoms, and then gastrointestinal symptoms such nausea, vomiting, diarrhea and abdominal pain. Huang Hanping. et al[30] analyzed the clinical characteristics of 305 patients in the early stage of the pandemic

in Wuhan Jinyintan Hospital, China. They found that compared with symptoms in the early stages of disease, as the time interval lengthened, the incidence of cardiopulmonary symptoms increased significantly. A similar pattern was also found in the work of Barak Mizrahi. et al. [12].These results reflected that longer intervals may mean a higher likelihood of cardiopulmonary symptoms such as dyspnea, and systemic symptoms and/or gastrointestinal symptoms, which were exactly the group 4 symptoms in this article. This revealed that the occurrence of these three symptoms do have chronological order. The occurrence of group 4 was later than that of group 2 and 3.

Another concern was that the odds of symptom group 4 was higher in patients aged 45–64 years than aged ≥ 65 years. Despite of the immunosenescence and inflamm-aging[31], elderly people were not as likely to show more severe initially symptoms as expected. The influence of symptoms progression may could not be neglected. Results in this article showed that people aged 13–44 years and 45–64 years have more cases with longer time intervals, indicating a time delay for medical treatment in this population. Similar to this, a study of 14,168 hospitalized infected people in Belgium also found that working age group (aged 20–60 years) had longer intervals between symptoms onset and their visits to a doctor[32]. In addition, it observed that for elderly people in nursing homes there was shorter delay times. And among the subjects in our study, the prevalence of underlying disease in people with ≥ 65 was higher than that in younger people. Therefore, a possibility was that for the elderly people, out of the care for the underlying disease, abnormal body signals in this group may be more likely to be detected by caregivers or to be detected in medical facilities thanks to they were themselves in health care facilities when the pandemic occurred. In China, for example, China's national office for the elderly had also stressed the care services for the elderly during the epidemic[33]. In contrast, the ages 13–44 years and 45–64 years are exactly working ages, so this group may be more likely unable to see a doctor in time due to work-related concerns. The above may account for the fact that middle-aged people were more likely to have longer time delay for medical visit than older people, and as a result, had more severe symptoms when first diagnosed. Thus, considering the delayed effect, this article suggested that middle-aged people, from the perspective of early treatment and early detection, may be a group needing attention in the prevention and control of the epidemic. Measures such as publicity may can be taken to improve the timeliness of medical treatment for the working-age population. Besides, for employers, it may be possible to relieve people's work-related worries through the provision of labor insurance benefits.

In contrast to many published studies that mainly described only individual symptoms, this article focused on the associations between different symptoms, had explored symptoms cluster patterns. Based on it, we analyzed the population characteristics of each symptom group, and the influence of symptoms development over time was also considered. We found that except the asymptomatic group, the highest frequency of symptoms combinations in the three groups with obvious symptoms were only fever, only expectoration and fatigue with fever, respectively. This can promote our knowledge of the symptoms of COVID-19 and provide some information for identifying infections. In addition, we found there was a chronological order in the occurrence of three symptom groups, symptom group 4 occurred later than symptom group 2 and 3. Else, this article revealed that people of working age were more likely to have a delay for medical treatment, as a result, with higher proportion of symptom group 4.

Our work had several limitations. Firstly, for comorbidities, information such as severity and duration was not collected, so the influence effect of comorbidities may be biased by the heterogeneity of severity grade and duration of diseases. Secondly, for the time interval between symptom onset and the first visit, there may be information bias of the self-reported time of symptom occurrence. In addition, the cluster patterns found in this article was just a probable result, more evidence are still needed in the future.

Conclusions

We creatively focused on and discovered different cluster patterns of COVID-19 symptoms. Under each pattern the common symptom combinations were revealed. It may be useful for identification of infected persons and distinguishing from other epidemic diseases at early stage. And we found that the middle-aged population may be a group requiring more attention during this epidemic, and some measures are expected to improve the timeliness of medical treatment for this group.

Declarations

Authors' contributions

XC, HW and TZ conceptualized the analysis. XC and HW implemented statistical Analysis. XC, HW, HY, JZ and TZ contributed to the study implementation, interpretation of results, and writing of the manuscript. LZ, CX, SM, ZL, FH, CY, WZ did the data collection and cleaning. All authors reviewed and provided comments on the manuscript. All authors approved the final version.

Availability of data and materials

The datasets used and analyzed during this study available from the corresponding author on reasonable request. We confirmed that these patients have not been reported in any other submission by all authors or anyone else.

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Sichuan Provincial Center for Disease Control and Prevention (NO. SCCDCIRB 2020-007). All methods were carried out in accordance with relevant guidelines and regulations. Written informed consent was obtained from each of the subjects.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. World Health Organization. Coronavirus disease 2019 (COVID-19) situation report 2020. 2020; [https://covid19.who.int/]

2. Furukawa K, Arii J, Nishimura M, Tjan LH, Lystia Poetranto A, Ren Z, Aktar S, Huang JR, Sutandhio S, Kurahashi Y et al. Seroepidemiological Survey of the Antibody for Severe Acute Respiratory Syndrome Coronavirus 2 with Neutralizing Activity at Hospitals: A Cross-sectional Study in Hyogo Prefecture, Japan. Jma j 2021; 4(1):41–49.

3. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med 2020; 382(18):1708–1720.

4. Eythorsson E, Helgason D, Ingvarsson RF, Bjornsson HK, Olafsdottir LB, Bjarnadottir V, Runolfsdottir HL, Bjarnadottir S, Agustsson AS, Oskarsdottir K et al. Clinical spectrum of coronavirus disease 2019 in Iceland: population based cohort study. Bmj 2020; 371:m4529.

5. Lv Z, Lv S. Clinical characteristics and analysis of risk factors for disease progression of COVID-19: A retrospective Cohort Study. Int J Biol Sci 2021; 17(1):1–7.

6. Young BE, Ong SWX, Kalimuddin S, Low JG, Tan SY, Loh J, Ng OT, Marimuthu K, Ang LW, Mak TM et al. Epidemiologic Features and Clinical Course of Patients Infected With SARS-CoV-2 in Singapore. Jama 2020; 323(15):1488–1494.

7. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. Jama 2020; 323(11):1061–1069.

8. Fried JA, Ramasubbu K, Bhatt R, Topkara VK, Clerkin KJ, Horn E, Rabbani L, Brodie D, Jain SS, Kirtane AJ et al. The Variety of Cardiovascular Presentations of COVID-19. Circulation 2020; 141(23):1930–1936.

9. Kariyawasam JC, Jayarajah U, Riza R, Abeysuriya V, Seneviratne SL. Gastrointestinal manifestations in COVID-19. Transactions of The Royal Society of Tropical Medicine and Hygiene 2021.

10. Diaz-Guimaraens B, Dominguez-Santas M, Suarez-Valle A, Pindado-Ortega C, Selda-Enriquez G, Bea-Ardebol S, Fernandez-Nieto D. Petechial Skin Rash Associated With Severe Acute Respiratory Syndrome Coronavirus 2 Infection. JAMA Dermatol 2020; 156(7):820–822.

11. Gautier JF, Ravussin Y. A New Symptom of COVID-19: Loss of Taste and Smell. Obesity (Silver Spring) 2020; 28(5):848.

12. Mizrahi B, Shilo S, Rossman H, Kalkstein N, Marcus K, Barer Y, Keshet A, Shamir-Stein N, Shalev V, Zohar AE et al. Longitudinal symptom dynamics of COVID-19 infection. Nat Commun 2020; 11(1):6208.

13. Peiris JS, Chu CM, Cheng VC, Chan KS, Hung IF, Poon LL, Law KI, Tang BS, Hon TY, Chan CS et al. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. Lancet 2003; 361(9371):1767–1772.

14. Zhao Y, Sun L, Bouchard HC, Zhang XX, Wan G, Hao YW, He SX, Jiang YY, Pang L. Coronavirus Disease 2019 versus Influenza A in Children: An Observational Control Study in China. Biomed Environ Sci 2020; 33(8):614–619.

15. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, Liu S, Zhao P, Liu H, Zhu L et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir Med 2020; 8(4):420–422.

16. Bohn MK, Hall A, Sepiashvili L, Jung B, Steele S, Adeli K. Pathophysiology of COVID-19: Mechanisms Underlying Disease Severity and Progression. Physiology (Bethesda) 2020; 35(5):288–301.

17. Tian S, Hu N, Lou J, Chen K, Kang X, Xiang Z, Chen H, Wang D, Liu N, Liu D et al. Characteristics of COVID-19 infection in Beijing. J Infect 2020; 80(4):401–406.

18. Eythorsson E, Helgason D, Ingvarsson RF, Bjornsson HK, Olafsdottir LB, Bjarnadottir V, Runolfsdottir HL, Bjarnadottir S, Agustsson AS, Oskarsdottir K et al. Clinical spectrum of coronavirus disease 2019 in Iceland: population based cohort study. (1756–1833 (Electronic)).

19. Guan WA-OX, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC et al. Clinical Characteristics of Coronavirus Disease 2019 in China. (1533–4406 (Electronic)).

20. Zhao Y, Sun L, Bouchard HC, Zhang XX, Wan G, Hao YW, He SX, Jiang YY, Pang L. Coronavirus Disease 2019 versus Influenza A in Children: An Observational Control Study in China. (2214 – 0190 (Electronic)).

21. Zhang J, Ding D, Huang X, Zhang J, Chen D, Fu P, Shi Y, Xu W, Tao Z. Differentiation of COVID-19 from seasonal influenza: A multicenter comparative study. Journal of medical virology 2021; 93(3):1512–1519.

22. Faury H, Courboulès C, Payen M, Jary A, Hausfater P, Luyt C, Dres M, Pourcher V, Abdi B, Wirden M et al. Medical features of COVID-19 and influenza infection: A comparative study in Paris, France. The Journal of infection 2021; 82(2):e36-e39.

23. Li P, Wang Y, Peppelenbosch MP, Ma Z, Pan Q. Systematically comparing COVID-19 with the 2009 influenza pandemic for hospitalized patients. International journal of infectious diseases: IJID : official publication of the International Society for Infectious Diseases 2021; 102:375–380.

24. Cobb NL, Sathe NA, Duan KI, Seitz KP, Thau MR, Sung CC, Morrell ED, Mikacenic C, Kim HN, Liles WC et al. Comparison of Clinical Features and Outcomes in Critically Ill Patients Hospitalized with COVID-19 versus Influenza. Annals of the American Thoracic Society 2021; 18(4):632–640.

25. Auvinen R, Nohynek H, Syrjänen R, Ollgren J, Kerttula T, Mäntylä J, Ikonen N, Loginov R, Haveri A, Kurkela S et al. Comparison of the clinical characteristics and outcomes of hospitalized adult COVID-19 and influenza patients - a prospective observational study. Infectious diseases (London, England) 2021; 53(2):111–121.

26. Coronavirus Disease 2019 in Children - United States, February 12-April 2, 2020. MMWR Morb Mortal Wkly Rep 2020; 69(14):422–426.

27. Zimmermann P, Curtis N. Why is COVID-19 less severe in children? A review of the proposed mechanisms underlying the age-related difference in severity of SARS-CoV-2 infections. Arch Dis Child 2020.

28. Papadopoulos VP, Koutroulos MV, Zikoudi DG, Bakola SA, Avramidou P, Touzlatzi N, Filippou DK. Diabetes-related acute metabolic emergencies in COVID-19 patients: a systematic review and meta-analysis. Diabetology international 2021:1–15.

29. Larsen JR, Martin MR, Martin JD, Kuhn P, Hicks JB. Modeling the Onset of Symptoms of COVID-19. Front Public Health 2020; 8:473.

30. Hung H XL, Zhang L, et al. Clinical characteristics of 305 COVID-19 patients in different stages in Jinyintan Hospital, Wuhan. Herald of Medicine 2020; 39(6):797–802.

31. Bajaj V, Gadi N, Spihlman AP, Wu SC, Choi CH, Moulton VR. Aging, Immunity, and COVID-19: How Age Influences the Host Immune Response to Coronavirus Infections? Front Physiol 2020; 11:571416.

32. Faes C, Abrams S, Van Beckhoven D, Meyfroidt G, Vlieghe E, Hens N. Time between Symptom Onset, Hospitalisation and Recovery or Death: Statistical Analysis of Belgian COVID-19 Patients. Int J Environ Res Public Health 2020; 17(20).

33. National Health Commission of the People's Republic of China. Notice of the National Office of Aging on Providing Care Services for the Elderly in the Normal Epidemic Prevention and Control. 2020; [http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=3726889dbf7f4cd0abd9a92105ae53ff]

Tables

Table 1 Characteristics of cases in different symptoms groups

Characteristics	Group 1 (n=541)	Group 2 (n=134)	Group 3 (n=106)	Group 4 (n=286)	Total (n=1067)
Age, median(P_{25},P_{75}),years	35(27,47)	42(28,50)	42.5(29.3,53.8)	43(33,51)	38(29,49)
Age group, n(%)					
0-12	26(65.00)	5(12.50)	6(15.00)	3(7.50)	40
13-44	349(56.93)	69(11.26)	50(8.16)	145(23.65)	613
45-64	140(40.70)	47(13.66)	41(11.92)	116(33.72)	344
≥65	26(37.14)	13(18.57)	10(14.29)	21(30.00)	70
Gender					
Male	388(55.51)	84(12.02)	61(8.73)	166(23.75)	699
Female	153(41.58)	50(13.59)	45(12.23)	120(32.61)	368
Hypertension					
Yes	23(31.51)	9(12.33)	10(13.70)	31(42.47)	73
No	511(51.83)	125(12.68)	95(9.63)	255(25.86)	986
Unidentified	7(87.50)	0(0)	1(12.50)	0(0)	8
Diabetes					
Yes	2(20.00)	3(30.00)	5(50.00)	0(0)	26
No	532(68.38)	131(16.84)	100(12.85)	15(1.93)	1033
Unidentified	7(2.51)	0(0)	1(0.36)	271(97.13)	8
Lung disease					
Yes	7(21.88)	8(25.00)	2(6.25)	15(46.88)	20
No	527(51.31)	126(12.27)	103(10.03)	271(26.39)	830
Unidentified	7(87.50)	0(0)	1(12.50)	0(0)	216
Cardiovascular disease					
Yes	5(20.00)	4(16.00)	2(8.00)	14(56.00)	22

No	529(51.16)	130(12.57)	103(9.96)	272(26.31)	1037
Unidentified	7(87.50)	0(0)	1(12.50)	0(0)	8
Imported cases					
Yes	354(80.45)	40(9.09)	23(5.23)	23(5.23)	440
No	187(29.82)	94(14.99)	83(13.24)	263(41.95)	627
Cluster cases					
Yes	146(51.77)	40(14.18)	34(12.06)	62(21.99)	282
No	324(48.80)	81(12.20)	58(8.73)	201(30.27)	664
Unidentified	71(58.68)	13(10.74)	14(11.57)	23(19.01)	121

Table 2 Results of factors associated with different symptom groups

Characteristics	Group 2		Group 3		Group 4	
	P	OR(95% CI)	P	OR(95% CI)	P	OR(95% CI)
Age(Referenced with 0-12 years age group)						
13-44 vs. 0-12	0.721	1.21(0.42-3.45)	0.552	0.74(0.27-2.03)	0.032**	4.08(1.13-14.76)
45-64 vs. 0-12	0.242	1.89(0.65-5.49)	0.628	1.29(0.46-3.61)	0.007	5.91(1.61-21.7)
≥65 vs. 0-12	0.543	1.49(0.41-5.44)	0.796	0.84(0.23-3.1)	0.288	2.22(0.51-9.7)
Age(Referenced with ≥65 years age group)						
0-12 vs. ≥65	0.543	0.67(0.18-2.44)	0.795	1.19(0.32-4.37)	0.287	0.45(0.1-1.96)
13-44 vs. ≥65	0.639	0.81(0.34-1.95)	0.783	0.87(0.33-2.29)	0.139	1.83(0.82-4.09)
45-64 vs. ≥65	0.595	1.26(0.53-3.01)	0.374	1.53(0.6-3.92)	0.015	2.66(1.21-5.85)
Gender						
Female vs. male	0.685	0.91(0.58-1.44)	0.883	0.96(0.58-1.59)	0.324	0.82(0.56-1.21)
Comorbidities						
Hypertension yes vs. no	0.777	0.87(0.34-2.23)	0.943	1.04(0.39-2.75)	0.538	1.27(0.59-2.73)
Diabetes yes vs. no	0.111	7.69(0.63-94.66)	0.004	29.43(3-288.66)	0.001	41.72(4.56-381.92)

Characteristics	Group 2		Group 3		Group 4	
	P	OR(95% CI)	P	OR(95% CI)	P	OR(95% CI)
Lung disease yes vs. no	0.216	2.04(0.66-6.34)	0.525	0.58(0.11-3.09)	0.451	1.49(0.53-4.23)
Cardiovascular disease yes vs. no	0.854	1.18(0.2-7.15)	0.321	2.28(0.45-11.65)	0.056	3.64(0.97-13.73)
Imported cases						
yes vs. no	<0.001	0.2(0.12-0.33)	<0.001	0.13(0.07-0.24)	<0.001	0.04(0.02-0.06)
Cluster cases						
yes vs. no	<0.001	0.59(0.36-0.95)	<0.001	0.58(0.34-0.98)	<0.001	0.27(0.17-0.41)

Figures
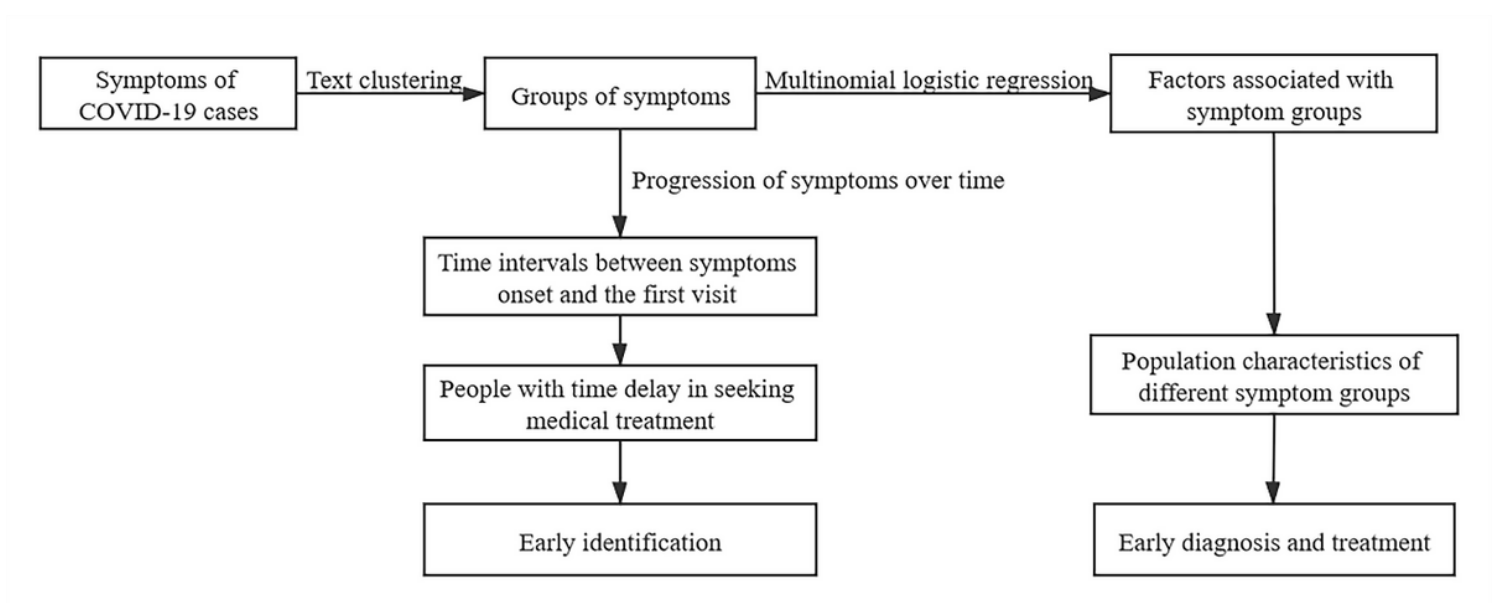
Figure 1
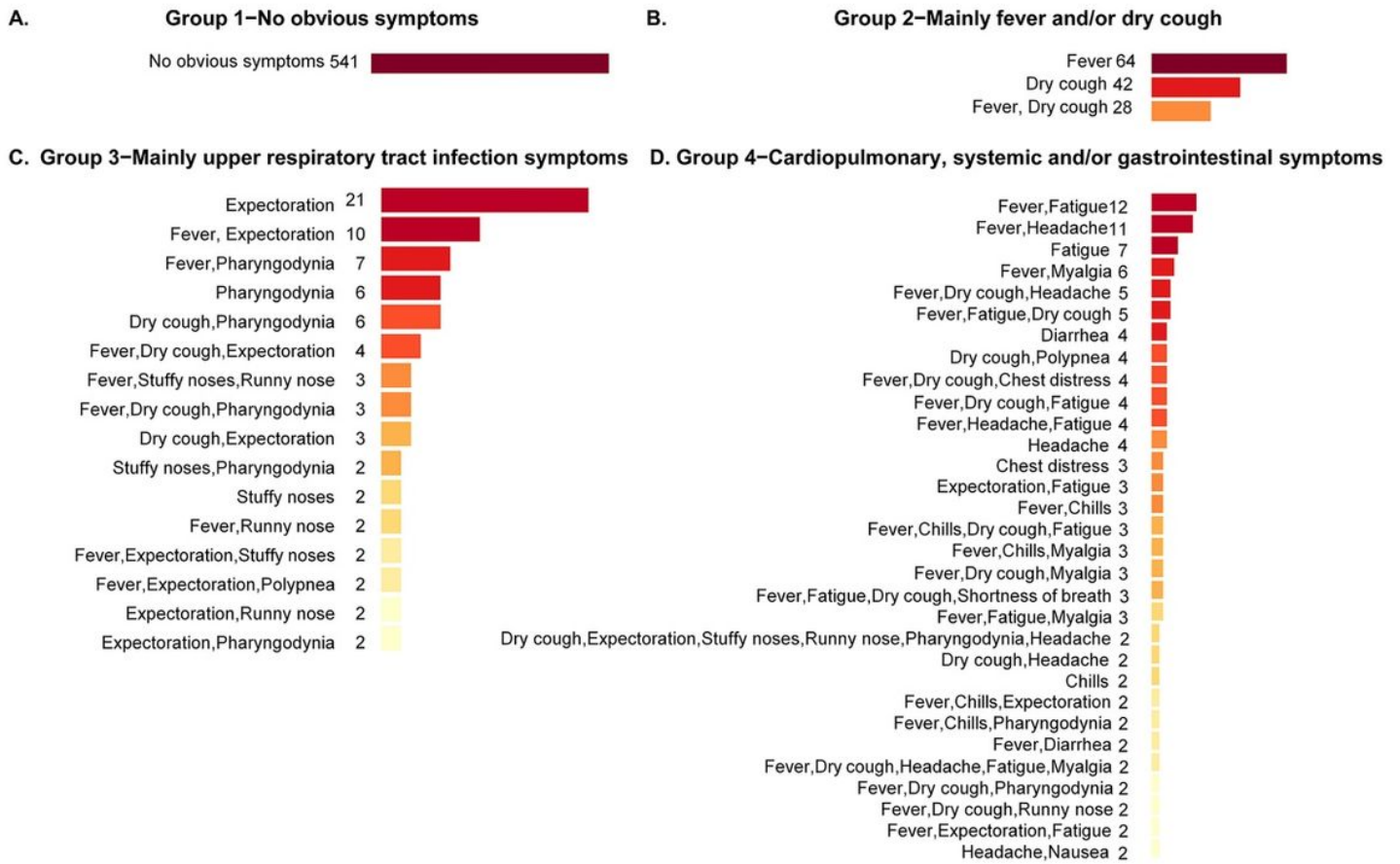
The procedure of analysis in this article.

Figure 2

Symptom combinations under different symptom groups. Symptom combinations with only one case in symptom groups 3 and 4 were not included.
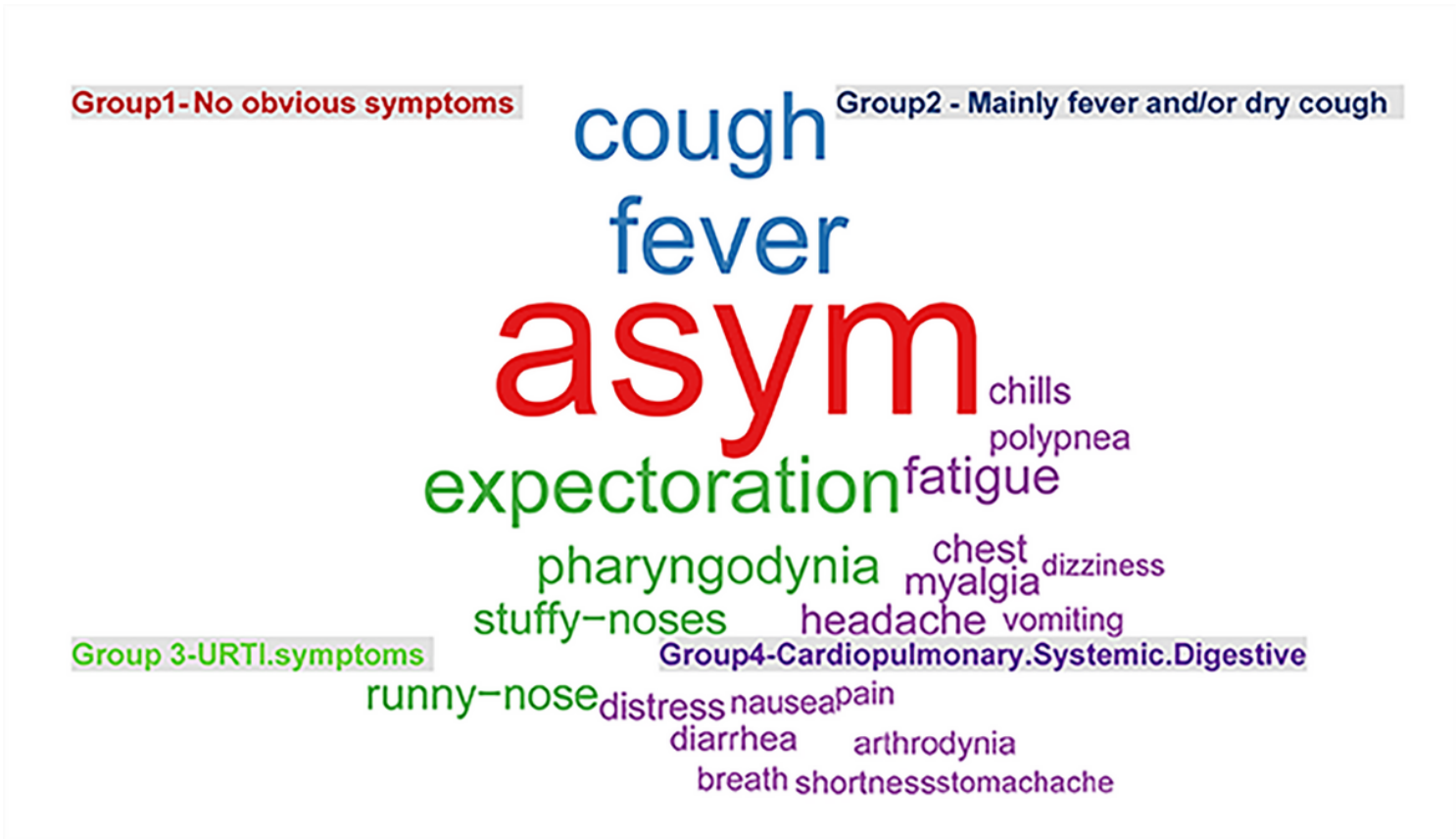
Figure 3

Symptoms word cloud under different groups. asym refers to no obvious symptoms ;URTI symptoms refers to mainly upper respiratory tract infection symptoms; Cardiopulmonary. Systemic. Gastrointestinal refers to cardiopulmonary, systemic and/or gastrointestinal symptoms.
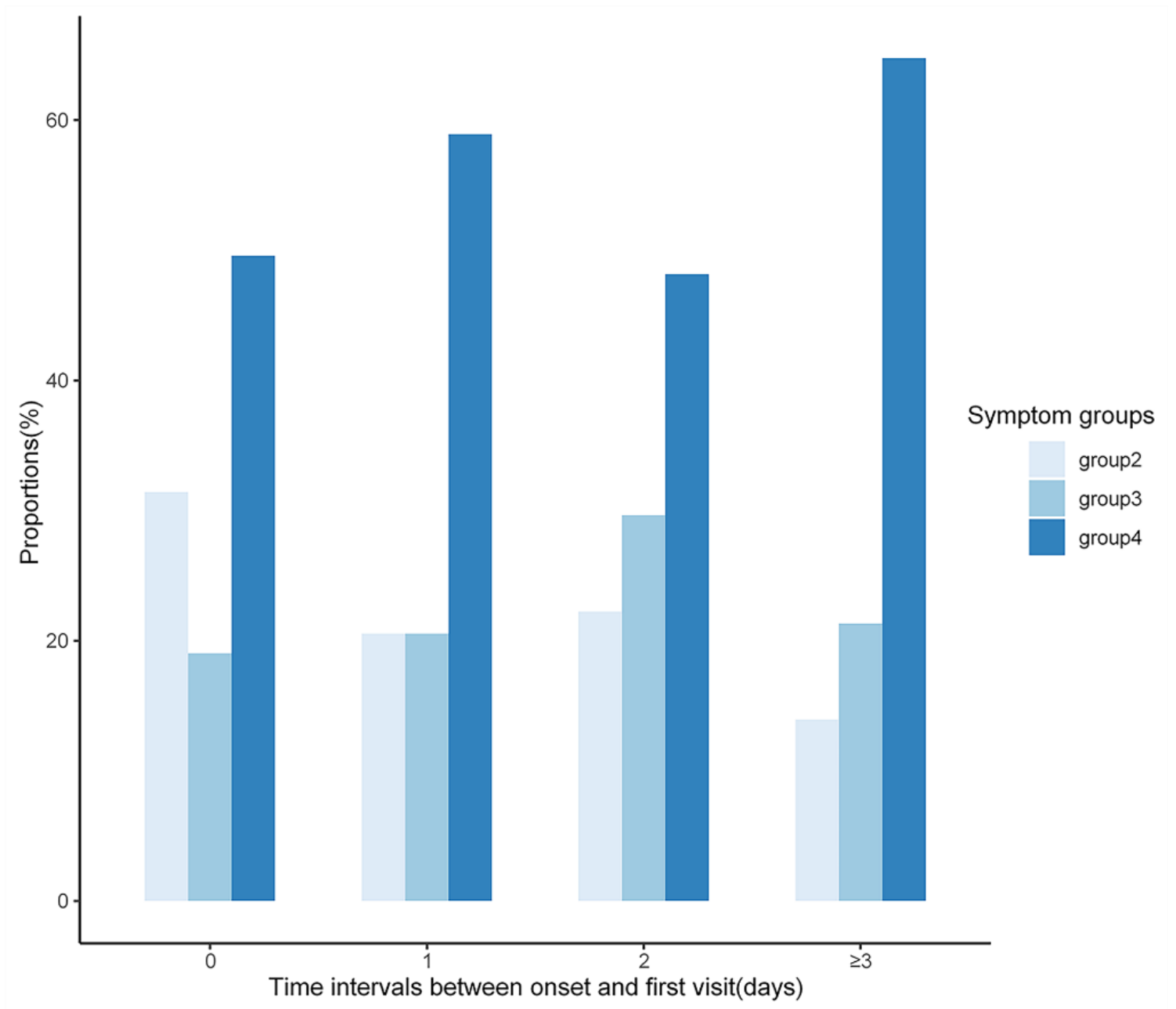
Figure 4

Time intervals between symptoms onset and the first visit in different symptom groups.
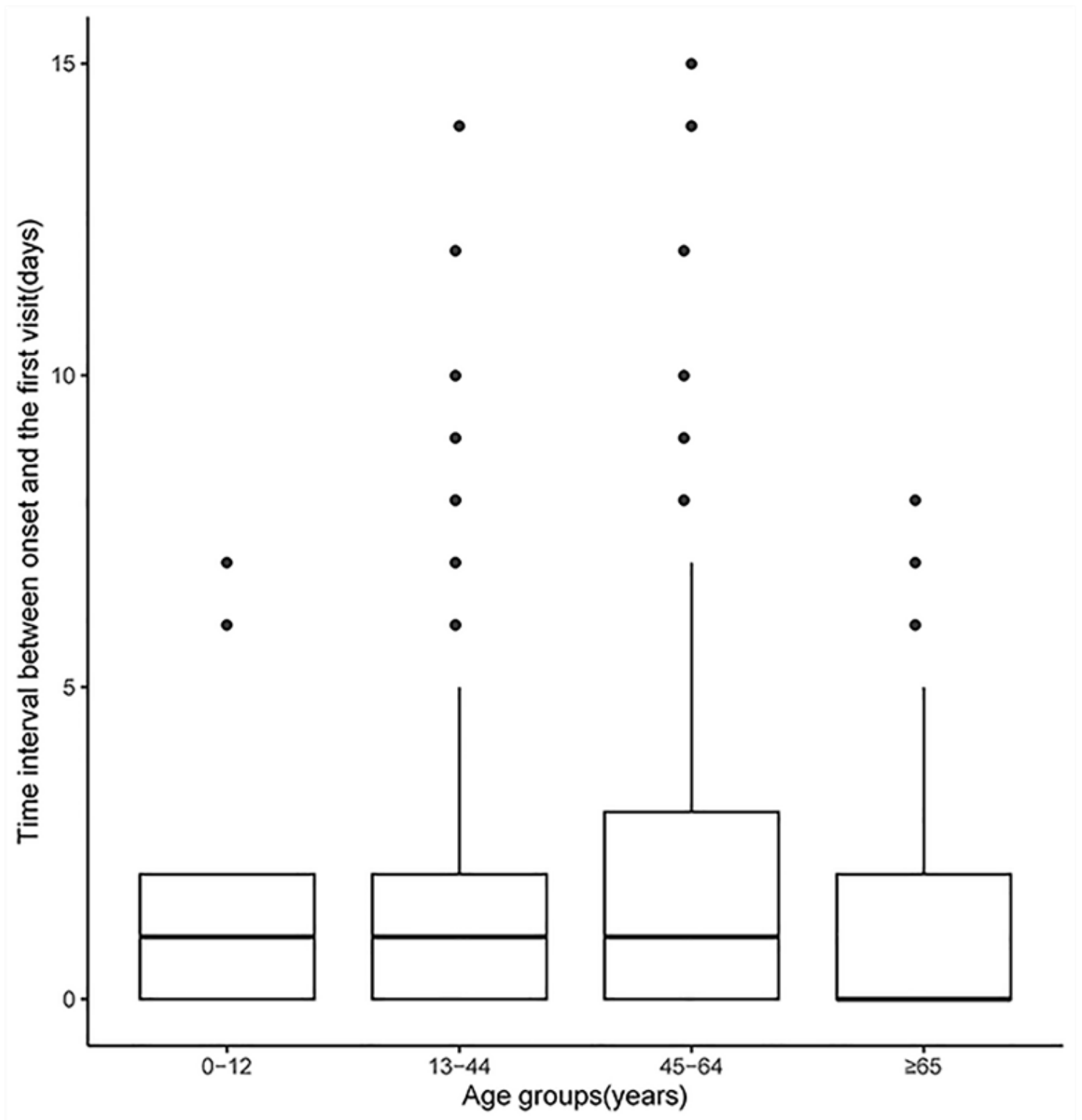
Figure 5

Time intervals between symptoms onset and the first visit in different age groups